

## **Historic, Archive Document**

Do not assume content reflects current scientific knowledge, policies, or practices.



# Probe

Newsletter for the USDA Plant Genome Research Program

Volume 1, No. 1/2

Spring/Summer 1991

## Strengthening Plant Genome Research Efforts—Goal of New USDA Program

*Jerome Miksche, Director  
Plant Genome Research Program, USDA*

**U**SDA's (U.S. Department of Agriculture) Plant Genome Research Program, established last October, will facilitate the improvement of plants—agronomic, horticultural, and forest species—by locating important genes and markers on chromosomes, determining the structure of those genes, and transferring the genes to improve performance. The end product will be superior plant varieties that more closely meet marketplace needs and niches, while creating a positive effect on the environment.

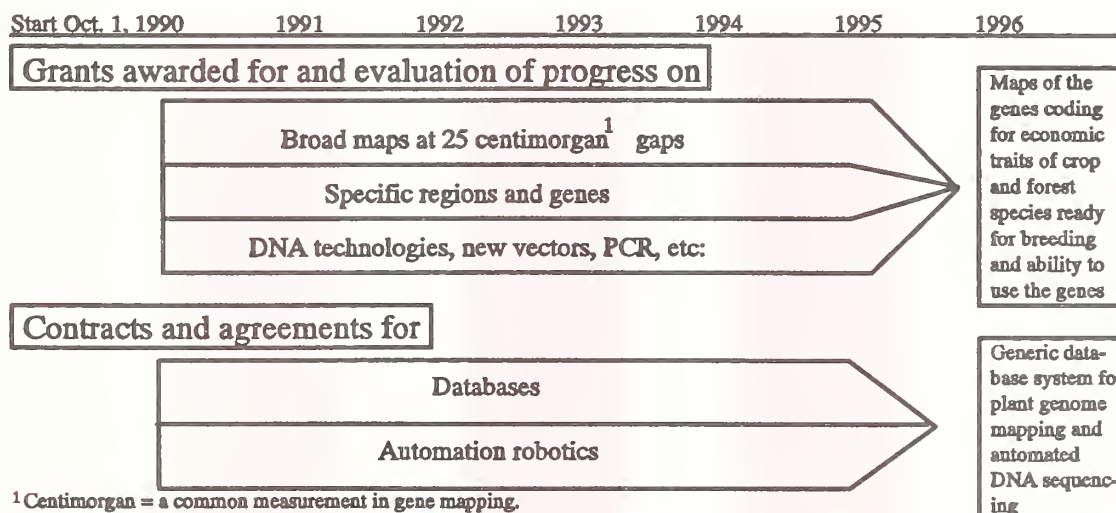
The program is a cooperative effort of several USDA agencies -- the Agricultural Research Service (ARS), the National Agricultural Library (NAL), the Cooperative State Research Service (CSRS), and the Forest Service (FS). ARS has the lead role in directing the program. With a budget of \$14.7 million for FY'91, the Plant Genome Research Program is managed through grants, contracts, and inter- and intra-agency transfers of funds.

### Competitive Grants

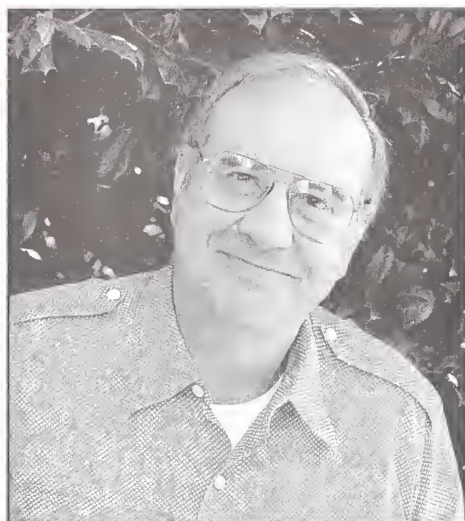
CSRS' Competitive Grants Research Office will manage the grants portion of the program in cooperation with ARS. Grants will be peer reviewed and mission oriented. Scientists from industry, academia, and government may

### Plant Genome Research Program

**Goal:** Facilitate the genetic improvement of plants (agronomic, horticultural, and forest species) by locating important genes and markers on chromosomes, determining the structure of those genes, and transferring the genes to improve performance to meet marketplace needs and niches.



<sup>1</sup>Centimorgan = a common measurement in gene mapping.



Dr. Jerome Miksche was named Director of the Plant Genome Research Program in 1989. Prior to joining the U.S. Department of Agriculture (USDA), Dr. Miksche was Head of the Botany Department at North Carolina State University. In 1985, he was named National Program Leader for Plant Physiology and Biotechnology within the Agricultural Research Service (ARS), which is a National Program Staff (NPS) appointment, the advisory body for ARS. Dr. Miksche has maintained his involvement with NPS while undertaking his role in the Plant Genome Research Program.

economic gain to American agriculture. Knowledge acquired from one commodity can be transferred to another crop species.

In this phase of the program, proposals anticipated will represent the following commodities: corn, soybean, tomato, wheat, barley, rice, pine (conifer), potato, garden bean, cotton, pea, peach, oat, sorghum, sweet potato, carrot, onion, apple, rose, sugarcane, citrus crops, and other agriculturally important species.

2) Grants will be awarded to develop more specific information on crops for which some data have already been acquired. Many of the major acreage crop species such as corn, soybean, wheat, and rice come under this category.

Scientists will determine gene construction relative to important specific traits such as yield, heat and cold tolerance, disease resistance, quality changes, drought tolerance,

gene transfer, and expression. This objective is important because it will generate results that offer products to the agricultural community. The analysis of factors in gene systems or gene families that regulate gene expression is required.

3) Progress in the Plant Genome Research Program is tied to developing new mapping and sequencing technologies. The following are examples of new technologies that need development: A new method to tag sequences, which will eliminate the need to store mapping probes, innovative applications of the polymerase chain reaction (PCR), new methodology to identify Quantitative Trait Loci (QTL's), creative computer software designed specifically for plant gene systems, *in situ*-hybridization technologies for plant chromosomes, methods that allow mapping of polyploid genomes, chromosomal sorting and separation

Cont. on page 21 ►

apply for grant funding.

Multidisciplinary submissions will be given favorable consideration.

The program grants consists of three components:

1) Support will be committed to constructing broad maps that locate important genes or gene systems in crops and forest species. This will be achieved using a technology that allows scientists to determine rather broad genetic similarities and differences, initiate assignment of DNA fragments on chromosomes, and then begin the mapping process. Budgeted dollars will not be allocated on a commodity basis, but on targeted gene systems or traits of some of the commodities that yield

## TABLE OF CONTENTS

Strengthening Plant Genome Research Efforts—Goal of New USDA Program.....	1
NAL's Plant Genome Center—A New Direction for Library Services .....	3
A Look at USDA's Competitive Grants Process .....	4
Plant Genome Database—Update.....	5
Planning Underway for USDA Soybean Genome Database.....	7
Forest Tree Genome Database—Development Begins .....	7
Information Superhighway Envisioned—Legislation Pending .....	8
French Join the International Human Genome Effort .....	10
Improving Access to Standardized Biological Terminology .....	11
Libraries Link Users With Specialized Databases .....	13
Theory and Application of YAC Technology for Genome Research .....	14
Parser Available for GenBank® Flat File.....	18
Genome Sequencing Conference III Set for September .....	18
Introducing Dr. Stephen Heller.....	19
Calendar of Upcoming Genome Events .....	20
Beltsville Symposium .....	22

## NAL's Plant Genome Center— A New Direction for Library Services

*Joseph H. Howard  
Director, National Agricultural Library  
U.S. Department of Agriculture*

When the National Agricultural Library (NAL) opened in Beltsville, Maryland, in 1967, all the technological advances in information management that would develop over the next 24 years were probably not envisioned by the staff.

But, since that time, NAL has worked diligently to stay abreast of the new technologies and to adapt those that most benefit users. These technological advances combined with the talents and dedication of library staff have enabled NAL to improve and expand its services to better meet the needs of users in the agricultural community.

Recently NAL opened the Plant Genome Data and Information Center (PGDIC), which offers a new direction for the library in archiving and in providing the public access to a national scientific database. An initiative of USDA's Plant Genome Research Program, the Center's primary goal is to make information on plant genome research readily available in useful formats to users.

This newsletter is but one example of the various services offered by PGDIC. In addition, PGDIC staff will provide information on all aspects of plant genome mapping; identify current genome mapping research; refer users to organizations or experts in the plant genome subject area; perform brief AGRICOLA (AGRICultural OnLine Access) database searches on a complimentary basis or exhaustive searches on a cost-recovery basis; furnish users with a Quick Bibliography (QB), Special Reference Brief (SRB), or user guide to literature; and assist users in accessing NAL's extensive collection.

These services<sup>1</sup> are available to anyone interested in plant genomes, including scientists, breeders, educators, students, legislators, information professionals, administrators, and the general public.

Plans are underway to develop a new database that will contain plant genome data on four agricultural commodities—soybean, corn, wheat, and pine. As producer of the



*Since 1983, Joseph H. Howard has served as Director of the National Agricultural Library (NAL). With 2 million volumes, NAL is one of the largest agricultural libraries in the world and is one of three national libraries in the United States.*

AGRICOLA computerized database and as the foremost agricultural library in the world, with nearly 2 million volumes and subscriptions to 26,000 periodicals, NAL is the ideal location for establishing and maintaining the database.

NAL staff anticipates that PGDIC will lead the way for additional centers and databases in other agriculturally related areas.

For more information on PGDIC, contact Susan McCarthy, coordinator for the Center, on (301) 344-3875. ♦

1. All programs of the USDA are available to anyone without regard to race, creed, sex, or national origin.

## Competitive Edge



# A Look at USDA's Competitive Grants Process

*Anne Datko, Program Director, National Research Initiative Competitive Grants Program, (NRICGP), CSRS, USDA*

## National Research Initiative

USDA has announced a national research initiative on agriculture, food, and the environment. The need for an increased investment in competitively awarded research was identified by the National Research Council of the National Academy of Sciences and broadly endorsed by the agricultural, scientific, and users communities. In the 1990 Farm Bill, the initiative was authorized at a full funding level of \$500 million per year. The Office of Management and Budget endorsed the full funding level and is committed to reaching this level by phased growth over the next few years. For FY'91, Congress funded the initiative at \$73 million; the President's FY'92 budget recommends funding at the level of \$125 million.

## NRICGP's Role in Supporting Plant Genome Research

In 1991, the National Research Initiative Competitive Grants Program subsumed the previously existing USDA Competitive Research Grants Program. Further, it was announced that while the Agricul-

tural Research Service (ARS) would be the lead agency for the USDA Plant Genome Research Project, the competitively awarded research grant component would be administered by CSRS and funded within the Program's Plant Systems Division at a level of \$11 million.

These research grants support research projects judged to further USDA programs. A solicitation announcing each program's guidelines is published in the Federal Register at the beginning of each fiscal year. Proposals for such awards may be submitted by any public or private institution, or individual.

## Peer Review

In general terms, peer review is used to provide the best possible scientific advice before expending Federal funds. Peer review is arranged by a program director -- a USDA scientist responsible for overseeing the review process, upholding the highest standards of conduct, and providing assistance and advice to the panel manager. A panel manager, an active researcher in the appropriate scientific community chosen each year, is

responsible, in consultation with the chief scientist of NRICGP, for selecting panel members with appropriate scientific expertise, review experience, and breadth of knowledge.

Written reviews are also solicited from the scientific community on an ad hoc basis. The peer review panel, chaired by the panel manager, considers the judgement of panel



*NRICGP anticipates publishing the 1992 solicitation in the Federal Register early in October 1991 with postmark deadlines for proposals being earlier than in 1991.*



members and the ad hoc reviews to provide advice to the NRICGP as to the scientific merit of each proposal. The program recommendation for funding, based on the panel ranking, is presented to the chief scientist who recommends the awards.

## Plant Genome Proposals

More specifically, the 1991 NRICGP solicitation identified research in the broad area of plant genome studies as having the potential to significantly improve agricultural and forest

## Home Base



productivity, and described the areas in which proposals would be accepted. In response to the solicitation, a large number of proposals were submitted by the postmark deadline of January 28, 1991. Dr. Maureen Hanson, director of the NSF<sup>1</sup>/DOE<sup>2</sup>/USDA Plant Science Center at Cornell University, was named panel manager for the Plant Genome Program for 1991, while Dr. Anne Datko, of the NRICGP staff, served as program director.

The peer review panel has already met to review proposals submitted; award recommendations have been made to Dr. Paul Stumpf, chief scientist of NRICGP. Because of his joint responsibility for NRICGP and the Plant Genome Research Program, Dr. Miksche was not only consulted as to panel member selection but also received the award recommendations. The grants funded will be published in January 1992 in the "Food and Agriculture Competitively Awarded Research and Education Grants" publication prepared by USDA's CSRS.

### 1992 Solicitation

NRICGP anticipates publishing the 1992 solicitation in the Federal Register early in October 1991 with the postmark deadline for proposals being earlier than in 1991. All institutions that received the 1991 solicitation and all 1991 applicants will

## Plant Genome Database—Update

*Douglas Bigwood, Database Manager  
Plant Genome Data and Information Center  
National Agricultural Library, USDA*

Providing users with fast, easy access to plant genome mapping and related information is a primary goal of USDA's Plant Genome Research Program. Currently plans are underway to develop a plant genome database system at NAL's Plant Genome Data and Information Center (PGDIC). The Center's database manager will direct the implementation of the plant genome database, which will contain public plant genome information for four

agricultural species—maize, soybean, wheat, and loblolly pine. In addition, procedures will be implemented to ensure that the information provided is up to date.

### Project Activities

Initial activities of the database project include site visits by PGDIC staff to several institutions also involved in developing genome information systems. Institutions visited include the National Center ►

automatically receive a copy of the 1992 solicitation. Other interested individuals may request the Solicitation (in September 1991, see the address below).

A vacancy announcement is expected to be issued within the next few months by NRICGP's Plant Systems Division for the position of Plant Genome Program Director. Persons interested in applying for this full-time, permanent USDA position may contact CSRS at the address below for additional information.

For solicitation information, contact the following:

Attn: 1992 Solicitation  
NRICGP/CSRS/USDA  
Room 323, Aerospace Building  
Washington, DC 20250-2200  
Phone (202) 401-5022 or  
FAX (202) 401-6480

For Genome Program Director vacancy information, write to:

Attn: Dr. Sally Rockey,  
Plant Systems Director  
NRICGP/CSRS/USDA  
Room 323, Aerospace Building  
Washington, DC 20250-2200  
Phone (202) 401-5114 or  
FAX (202) 401-6488 ♦

1. National Science Foundation

2. U.S. Department of Energy

for Biotechnology Information at the National Library of Medicine (GenInfo® Backbone), the Los Alamos National Laboratory (GenBank®), the Lawrence Berkeley Laboratory (Chromosome Information System), the Welch Library at Johns Hopkins University (Genome Data Base), the Massachusetts General Hospital (*Arabidopsis* mapping project), and Agrigenetics (commercial breeding projects). NAL staff benefited from the wealth of knowledge and experience provided by these groups. Hopefully, as a result of this information sharing, some of the pitfalls and problems faced by other institutions can be avoided in the USDA project.

The Center also has been active in two CODATA projects: Biological Macromolecules (seeking to improve coordination among institutions that compile protein and DNA sequence data) and Standardized Terminology for Access to Biological Data Banks (headed by Lois Blaine, whose article appears elsewhere in the newsletter). CODATA is an interdisciplinary scientific committee of the International Council of Scientific Unions that seeks to improve the quality,

reliability, management, and accessibility of data important to all fields of science and technology.

## Species Groups

The task of collecting and evaluating the data that will comprise the plant genome database system is the responsibility of the principal investigators for the four plant species and their advisory committees. The principal investigators are Frank Greene and Olin Anderson (wheat), David Neale (pine), Ed Coe (maize), and Randy Shoemaker (soybean). Each group will have its own database requirements. Cooperators in the project have made a concerted effort to ensure that all database-related activities are performed in a coordinated manner. The ultimate goal is to provide a master database design that is as generic as possible. If this goal is achieved—and efforts so far are encouraging—data from a number of additional species may be easily incorporated in the database in the future. Furthermore, plans are to develop an open system so USDA's database can forge data links with

related data sources such as GenBank®, AGRICOLA, and the Germplasm Resources Information Network (GRIN).

## Future Plans

By the time this newsletter is printed, the first meeting of the PGDIC Technical Committee will have been held. Composed of genetic and information experts, the committee is expected to be valuable in ensuring that the plant genome database is the best possible resource for users.

PGDIC staff are also establishing computer and communication systems. Initial development will be performed on Unix<sup>1</sup> workstations using the Sybase<sup>1</sup> relational database management system. These have essentially become de facto standards in the genome community. The database system's major network access will be through Internet via a T1 line.

The database analysis and design are proceeding as planned. Implementation will begin in the near future. ♦

1. Mention of a trade name or brand name does not constitute endorsement or recommendation by the Department over similar products not named.

## DESERT STORM RAINS ON PLANT GENOME GRANT PROPOSALS

January 28, 1991, was the deadline for plant genome grant proposals. Prior to the Persian Gulf crisis, express mail was delivered directly to the Aerospace Building. High-level security protocols were engaged with the commencement of hostilities in the Gulf. The express mail was subsequently routed to the main USDA mailroom two blocks away and X-rayed. A few grant proposals were inadvertently returned to the senders. All of these proposals were subsequently accepted for consideration.

*Touching Base with Randy Shoemaker*



## Planning Underway for USDA Soybean Genome Database

*Randy C. Shoemaker  
Research Geneticist, Field Crops Research Unit  
Iowa State University  
Agricultural Research Service, USDA*

Until recently, information on soybean genetics has developed slowly in relation to data gathered on other major crops. However, the increased speed with which genetic data on the soybean currently is being accumulated rivals that of any other genetics system.

The importance of the soybean as a major world oilseed crop plus the increased volume of genetic information accumulating have made the soybean an important focus of the USDA Plant Genome Research Program's thrust to develop a plant genome database management system. The database, to be located at the National Agricultural Library (NAL), will include information on four agricultural commodities—soybean, corn, wheat, and pine.

### **Soybean Conference Held**

At a conference held recently in St. Louis, Missouri, over 30 participants from 14 States and Canada met to discuss developing a prototype soybean genome database. Participants provided information on the long-term needs of the project

and established a priority for accomplishing the tasks required to develop a database.

The group included scientists from State and Federal institutions and private industry, and representatives from the Germplasm Resource Information Network (Mark Bohning), the American Soybean Association (Keith Smith), the USDA Plant Genome Research Program (Jerry Miksche), and the National

Agricultural Library (Susan McCarthy). In addition, Mary Berlyn, a cooperator in the USDA Program and co-developer and curator of the *E. coli* genetic stock center database at Yale University, provided useful input and facilitated discussions.

### **Working Committees Formed**

Seven committees listed below were formed during the conference to address specific concerns according to their areas of expertise. The committees will determine the probable relationships between their respective segment of the database and the other

*Cont. on page 9 ►*

*Touching Base with David Neale*



## Forest Tree Genome Database—Development Begins

*David Neale, Molecular Geneticist  
Institute of Forest Genetics (IFG)  
USDA Forest Service, Berkeley, CA*

Plans are underway by USDA's Forest Service (FS) staff to develop a prototype genome mapping database for the loblolly pine, an important forest tree species. The database

project is part of USDA's Plant Genome Research Program effort to provide users with fast, easy access to plant genome data. USDA plans are to develop a plant genome database system at NAL's Plant Genome Data and Information Center, which will contain plant genome information on the loblolly pine and three additional

*Cont. on page 9 ►*

## From the Hill



# Information Superhighway Envisioned—Legislation Pending to Establish National Computer Network

*Susan McCarthy, Coordinator  
Plant Genome Data and Information Center  
National Agricultural Library, USDA*

A national superhighway for information may soon become a reality if Congress passes the proposed legislation needed to establish the National Research and Education Network (NREN)—a high-capacity, high-quality computer network that supports a broad set of applications and network services for the research and education community.

NREN would expand and upgrade the existing interconnected array of primarily scientific research networks that comprise Internet, including the nationwide NSFNET (the backbone), regional networks such as NYSERNET and SURANET, and local area networks. NSFNET, perhaps the best known of the Internet networks, allows researchers and educators to exchange up to 1.5 million bits of data per second. The proposed NREN is expected to be at least a thousand times faster.

**Facilitating Genome Research**  
Fast, high-quality networks (gigabit per second transmission rate) are

needed to facilitate access to genome data. NREN would link libraries, government research laboratories, industry, and universities. The National Agricultural Library and the National Library of Medicine are cited in the proposed legislation as focal points in the information distribution networks. These libraries play vital roles in the genome programs for humans, plants, and animals.

## Proposed Legislation

Senator Albert Gore, Jr. (D-TN) introduced a bill (S272) in the Senate this year to establish NREN under the High Performance Computing Act of 1991. Recently, Representative George Brown (D-CA) introduced a companion bill (HR656) in the House. Legislation was first proposed by Senator Gore in 1988. Last year a revised version of the bill was unanimously passed by the Senate, but the House failed to act on the companion bill. The current Senate and House bills have been placed on

the congressional calendar.

Under the proposed legislation, the National Science Foundation would provide leadership in establishing the new fiber-optic computer network in cooperation with the Department of Defense, the Department of Energy, the Department of Commerce, the National Aeronautics and Space Administration, and other agencies.

A third bill introduced this year proposes to establish a Federal High-Performance Computer Network, which would serve many of the same purposes envisioned for NREN. Senator J. Bennett Johnston (D-LA) is sponsoring the bill—the Department of Energy High-Performance Computing Act of 1991 (S343). The Department of Energy is designated lead agency under the proposed legislation.

## Increased Funds and Support

A new Presidential Initiative, "Grand Challenges: High-Performance Computing and Communications," issued by the Office of Science and Technology Policy calls for a 30-percent increase in funding for FY'92. The funds will support high-performance computing systems, advanced

## *Superhighway—cont. from page 8*

software technology and algorithms, NREN, basic research, and human resources.

Japan and Europe are well ahead of the United States in recognizing the need for an information infrastructure. Maintaining the United States' technological lead and competitiveness targets this critical technology for congressional action. Support of the proposed legislation will have a positive benefit for genome research programs. The supernetwork—expanded, upgraded, and connected—will maximize the benefits and technology-transfer opportunities derived from the genome projects. ♦

## *Soybean—cont. from page 7*

areas. In addition, the soybean team will examine strategies developed for other databases and genome initiatives, including those for human and bacterial genetics. The shaded box at left has the list of the committees and respective chairpersons. Group members have had extensive discussions with industry representatives to ensure that a maximal amount of user needs will be met. In addition, to avoid duplicated efforts, the Quantitative Traits Committee will interact closely with a similar committee of the maize database group. The Quality Control Committee will explore methods to maintain the integrity of the information in the database while facilitating access by its users. The soybean team also will work closely with the Germplasm Resource Information Network (GRIN) so that a smooth interface is established between the

## *Forest Tree—cont. from page 7*

species—corn, soybean, and wheat.

Work on the loblolly pine database project began in late May at the Institute of Forest Genetics (IFG), Pacific Southwest Research Station, in Berkeley, California. Several goals have been identified for FY'91. First, a schema must be determined for the database. IFG plans to collaborate with the Human Genome Computer Science Group at the Lawrence Berkeley Lab (LBL). A computer scientist is being recruited by IFG to work with the LBL group. Two other database projects to be completed in conjunction with LBL also have been identified: (1) development of an

electronic laboratory notebook for tree genome mapping and (2) image processing and analysis software for tree mapping data. In addition, IFG will collaborate with the University of Montana to develop statistical approaches and computational methods to map quantitative trait loci (QTL) in segregating tree pedigrees.

David Neale is the principal investigator for the loblolly pine database. A Forest Tree Genome Database Advisory Group will be established this summer. A workshop is planned for late 1991. ♦

modified GRIN database and the developing soybean database.

If the conference was any indication, the enthusiastic grass-

roots support shown there will assure the momentum needed to establish a comprehensive plant genome database for soybeans. ♦

### **Disease/Pathology:**

Roger Boerma, University of Georgia  
Phone (404) 542-0927

### **Germplasm:**

Jim Specht, University of Nebraska  
Phone (402) 472-1536

### **Maps:**

Nevin Young, University of Minnesota  
Phone (612) 625-2225

### **Metabolic Pathways:**

Tom Cheesbrough, South Dakota State University  
Phone (605) 688-5504

### **Organelles:**

Beth Grabau, Virginia Polytechnic Institute and State University  
Phone (703) 231-9597

### **Quantitative Traits and Quality Control:**

Randy Shoemaker, Iowa State University  
Phone (515) 294-6233

## Other pursuits



# French Join the International Human Genome Effort

Susan McCarthy, Coordinator  
Plant Genome Data and Information Center  
National Agricultural Library, USDA



Acting Director Dr. Jacques Hanoune of the French Human Genome Program met with Dr. Jerome Miksche, Director, USDA Plant Genome Research Program. Pictured above left to right: Dr. Michele Durand, Scientific Attache; Pierre Oudet, Acting Director of Informatics; Michel Cohen-Solal, Research Director at INSERM, and Dr. Jacques Hanoune.

The French government has made a major investment in bioscience research. They have established a new agency, the Groupement d'Intérêt Public (GIP), which will play a lead role in coordinating genome activities in France and with other countries.

GIP Acting Director Dr. Jacques Hanoune and Acting Associate Director for Informatics Dr. Pierre Oudet visited the United States this summer to study the organization of genome programs and corresponding

database developments for these programs. During their trip, they met with Dr. Jerome Miksche, Director of the USDA Plant Genome Research Program, and NAL staff. This visit introduced GIP representatives to U.S. genome efforts and provided an opportunity to discuss areas for future collaboration between the United States and France.

The French program will have three main scientific thrusts. 1) The primary effort will be mapping and sequencing the genes expressed in

man. Largely, this will include the sequencing of cDNA libraries. 2) The program will also support high resolution mapping of identified genes to chromosome coordinates, particularly those related to genetic diseases. 3) Model organisms will be studied to develop new and more efficient technologies and to understand gene function. The organisms under study include yeast, mouse, bacteria, and wheat.

A significant effort will involve the storage and retrieval of the mapping data, informatics. The high volume of data generated from the mapping efforts will require two basic developments: (1) advances in adapted informatic structures—in other words, user needs assessments—and (2) programs for the acquisition and analysis of the mapping and sequencing data. Approximately 20-25 percent of program funds will support informatics.

FY'91 will see a French investment of 50 million francs—about \$10 million—in new funds. This is added to the figure of approximately 150 million francs supporting existing genome programs. The budget for the next fiscal year is expected to reach 100 million francs in new funds. ♦

## Connections



# Improving Access Standardized Biological Terminology

*Lois Blaine*

*Head, Bioinformatics Department  
American Type Culture Collection  
Rockville, Maryland*

## CODATA Convenes Workshop To Address Problems

Formulating a plan to improve access to standardized terminology for biological database producers and users was the goal of a workshop held May 14-16 in Nancy, France, by the CODATA Commission on Standardized Terminology for Access to Biological Data.

The workshop, jointly sponsored by the U.S. National Center for Biotechnology Information and Commission of the European Communities DG XII, was attended by representatives of the Biological Unions of the International Council of Scientific Unions (ICSU), producers of bibliographic and factual databases, and professional terminologists. This combination of participants, coming from disparate subdisciplines of biological and information science, provided an excellent blend of appropriate talents to address the multifaceted problems of standardizing terminology.

A primary goal of the Commission, re-emphasized during the Nancy Workshop, is to raise the level

of consciousness within the biological community of the need to communicate across disciplines. A major benefit of today's computer technology is that it provides the means to integrate data in ways that will lead to new scientific insights. Artificial intelligence, innovative programming, massive data storage capabilities, and vastly improved communication technology will inevitably draw diverse data sources together. If data is to be integrated, exchanged, and searched efficiently, the intellectual input to make this possible must come from biologists now.

Although problems surrounding the "standardization" of nomenclature and terminology have been with us for centuries, information technology demands that we take a fresh look at these problems and devise new methods to solve them that takes full advantage of today's technological tools.

## Workshop Segments

The workshop program consisted of three segments. The first segment included of formal presentations that provided a background and an overview of the perceived problems in interdisciplinary access to biological terminology. Dr. Andrzej Elzanowski, from the German Max-Planck-Institut für Biochemie,

articulately summarized the problems faced by database producers in "translating" the nomenclature and terminology used by authors who write for scientific journals into some type of "standard" that can be used for consistency of retrieval of identical concepts.

During the workshop, it became apparent that authors, editors, publishers, and database producers all face a similar situation—the lack of clear guidelines on nomenclature and taxonomy of organisms, and the terminology to describe their characteristics.

In the second segment, representatives of the Biounions, in a series of roundtable discussions presented the "state of the art" within each union regarding nomenclature and terminology standards. Broad topics covered by the union representatives included botanical and zoological nomenclature and taxonomy, biochemistry, microbiology, pharmacology, physiology, nutrition, food science, and clinical medicine. Participants discovered that the nomenclature and terminology committees of the ICSU unions face as many problems in providing standards as database producers and users do in locating standardized terminology for biological concepts. ►

Their ability to provide wide access to standardized terminology in formats and electronic media desired by database producers is limited by the fact that almost all of their work is performed on a "volunteer basis" and is primarily designed for intradisciplinary use. Additional resources would be required to expand the work of the Biounion nomenclature and terminology committees.

During the third segment, participants heard from terminological specialists who discussed existing standards for terminology. There are general principles that must be applied when developing terminological databases, regardless of the scope and content. The Commission agreed to work with the specialists to educate biologists on the ICSU committees in the implementation of these principles. Several documents recognized by the International Standards Organization were recommended for use in the educational campaign.

## Increasing Information Exchange

The workshop discussions raised the participants' awareness that the present interdisciplinary nature of many scientific activities leads to a greater need for an exchange of information within the component parts. The impact of multinational projects, such as HUGO (Human Genome Organization), similarly imposes added demands for clarity and standardization of expression. The integration of international, interdisciplinary databases will require some precision in defining

terminology for uniform interpretation of scientific principles.

Workshop participants agreed that the first step in providing wider access to standardized terminologic references is to expand efforts initiated by the U.S. National Library of Medicine (NLM) in establishing a "Nomenclature File" in its Directory of Biotechnology Information Resources. All agreed that the NLM file is a useful beginning, but that a broader international inventory of terminological resources and their relationships to one another is required. A steering committee composed of Commission members will be established to design procedures for developing this international terminological inventory.

## "Term Bank" Needed

The ultimate goal of the Commission would be to catalyze the development of an international "Term Bank." This, of course, would have to be developed in modules and would necessitate preliminary studies to determine feasibility and user requirements. The Commission would also seek the cooperation of other international organizations such as the International Council on Scientific and Technical Information (ICSTI) and the International Federation of Scientific Editors (IFSE), both represented at the Nancy workshop. An enormous effort would be required to make such a term bank available to the international scientific community. While the computer and communication technology is available to link subsets of such a database, issues, including copyright,

cost recovery, coordination, updating responsibilities, and funding were recognized as potential barriers to the successful accomplishment of the goal.

Dr. Leslie Sobin, representative from the International Union Against Cancer (UICC), aptly set to verse the precautions that must be taken by the Commission in approaching a project of this size:

*When you're looking for the answer  
how to classify all cancer,  
proteins, microbes, fish and succulent  
legumes,  
You must know a little Latin  
tell a round fish from a flat one  
and have memory with lots and lots of  
room.  
But, before we start alinking  
we should sit back and be thinking  
on our methods, clientele and on our  
goal,  
Lest we make a mammoth bank  
rarely used and rarely thanked  
just consuming funds and efforts:  
A BLACK HOLE.*

—Leslie Sobin

Members of the CODATA Commission will keep these words of wisdom in the forefront as they launch their campaign to improve access to standardized biological terminology.

For further information on the CODATA Commission on Standardized Terminology for Access to Biological Data, please contact the CODATA Secretariat, 51 bd. de Montmorency, 75016 Paris, France. ♦

## Hard Copy



# Libraries Link Users With Specialized Databases

*Vincent Caccese, Librarian  
Biological & Agricultural Sciences Reference Department  
Shields Library  
University of California, Davis*

Some scientists, students, and others involved in research may not consider using library reference services when they need a database on molecular biology. They may fail to take advantage of a research librarian's access to nonbibliographic databases such as GenBank® or NBRF-PIR. But, today, an increasing number of research libraries have access to specialized databases to better serve users. Along with those mentioned, similarly recondite databases have recently become part of reference services. One example is molecular structure searching on chemical databases, such as Chemical Abstracts, which librarians in many universities and companies have successfully mastered to benefit their users.

## New Databases Created

Leading research libraries have recognized the importance of molecular biology databases to their mission. Increasingly, research libraries are not only using specialized databases but creating them as well. The National

Agricultural Library (NAL) and the National Library of Medicine (NLM), for example, are updating their bibliographic files (NAL's AGRICOLA and NLM's MEDLINE) to incorporate molecular biology data while creating new databases to integrate several kinds of reference/data files (NLM's GenInfo® Backbone is the most recent example.).

## University Library Services

Along with national libraries and special libraries in private industry, academic research libraries are important centers for genetics information. At Johns Hopkins University's Welch Medical Library, the database GDB/OMIM (Genome Data Base/Online Mendelian Inheritance in Man) has been available for some time from computers on public data lines to registered users around the country.

The Biological Sciences Library at Columbia University provides its students and faculty with access to locally mounted GenBank®. While the University of California, Davis, libraries do not use dedicated equipment for such access, users can

access sequence databases (such as GenBank®) mounted on the campus computer via general-purpose terminals in the libraries. With the additional availability of GenBank® and other sequence databases via SprintNet and Internet for daily updated information, it is likely that the user population will increase.

## User Assistance

Although libraries address local demands according to individual means and staffing, scientists and other users should not hesitate to ask for assistance in locating databases. Appropriate requests that do not make inordinate demands on staff or strain budgets include identifying a database, including its scope and currency; finding documentation for the database; identifying means of access and its cost; finding ancillary documentation in the scientific literature; and, if all else fails, suggesting who to call. Some libraries may also aid a new user in learning the system or in running a search. If the database requested is not available, the librarian will at least know that there is local demand for such information. ♦

## Theory and Application of YAC<sup>1</sup> Technology for Genome Research

Brian M. Hauge and Howard M. Goodman

Department of Genetics, Harvard Medical School

Department of Molecular Biology, Massachusetts General Hospital, Boston

A major challenge in plant molecular biology is isolating genes where the biochemical function of the gene product is unknown. In a variety of plant species, genes controlling a wide range of fundamental developmental and metabolic processes have been identified by mutational analysis and placed on classical genetic linkage maps. Examples include genes conferring resistance to plant pathogens, the synthesis and response to plant hormones, drought tolerance, and genes required for a variety of important developmental pathways. In most cases, while the mutant phenotype and genetic map locations are known, virtually nothing is known about the product of the gene.

### Gene-Cloning Methods

There are several ways to clone genes for which the genetic locus and not the product of the gene is known. If a gene can be tagged with a transposable element, the gene can be cloned directly by isolating the sequences flanking the site of insertion. The

cloning of genes by transposon tagging has been used extensively in maize. The most widely utilized and best characterized plant transposable elements, the maize Ac and Ds elements (reviewed by Fedoroff in Berg and Howe, 1989) are further, capable of transposing in various heterologous plants (Van Sluys et al,

### Developments

1987), thereby extending the utility of this system to plants having no well-characterized transposable element systems. In addition to endogenous transposons, the T-DNA of *Agrobacterium tumefaciens* has also been successfully used for gene tagging (Feldmann et al, 1989).

A second alternative is to clone genes corresponding to deletion mutations using the technique of genomic subtraction (Straus and Ausubel, 1990). This method is based on the progressive enrichment for DNA fragments present in the wild-type genome but absent in the mutant genome that harbors the

deletion. Following multiple rounds of enrichment, the resultant fragments are amplified by polymerase chain reaction (PCR) and cloned. The one constraint of the protocol is that the deletion must encompass a single restriction fragment which is composed entirely of non-repetitive DNA. Genomic subtraction has recently been used to clone the GA-1 gene from *Arabidopsis* (Sun and Ausubel, unpublished results).

Like the gene-tagging strategies, the advantage of genomic subtraction is that one has

immediate access to the gene(s) of interest. A disadvantage of both

approaches is that insertions or deletions in essential genes will be lethal, so phenotypes associated with "leaky" point mutations will not be detected.

Second, transposons as well as different mutagenic agents exhibit some degree of sequence specificity. Therefore, many important loci will be refractory to isolation by transposon tagging or genomic subtraction.

### Chromosome-Walking Strategy

A more general approach is to clone genes by chromosome walking. This strategy is general in the sense that the cloning of the gene is based solely on the mutant phenotype and genetic

map position. Therefore, chromosome walking can be used to clone any gene which can be genetically identified. The first step toward cloning the gene is to identify DNA probes residing within one to several cM (centimorgan) of the locus of interest. Typically this is achieved by analyzing the meiotic segregation of restriction fragment length polymorphisms (RFLPs). Once a linked RFLP(s) has been identified, it can be used as the starting point to initiate a chromosome walk.

Briefly, chromosome walking entails the progressive isolation and characterization of overlapping sets of genomic clones. The overlapping clones are selected by hybridization using end-specific probes (probes generated from the extremities of the clone/contig). The walk is continued in this manner until the region spanning the intervening gap has been bridged by an overlapping set of clones. While chromosome walking is technically straight forward, in practice the procedure is extremely labor-intensive and ill-suited for large projects where more than a few steps are required.

## Constructing Physical Maps

Recently, interest has focused on strategies for constructing physical maps of entire genomes. By definition, a physical map consists of a linearly ordered set of DNA fragments encompassing the genome or region of interest. Physical maps are of two types, macro-restriction maps and ordered clone maps. The former consists of an ordered set of large DNA fragments generated by using restriction enzymes whose recogni-

tion sequences are infrequently represented in the genome (Smith et al, 1986). The macro-restriction map provides information about the organization of DNA fragments at the level of the intact chromosome, thereby providing long-range continuity.

As the name implies, an ordered clone map consists of an overlapping collection of cloned DNA fragments. The DNA may be cloned into any one of the available vector systems—YACs, cosmids, phage, or even plasmids. Major advantages of ordered clone maps are that they are of high resolution and directly provide the clones for further study.

The immediate benefits of having a physical map are twofold. First, the physical map provides ready access to any region of the genome which can be genetically identified. Given a mutation of known genetic map location, the physical map can be used to easily isolate an overlapping collection of clones encompassing the locus of interest. By eliminating the need for labor-intensive steps such as chromosome walking, researchers are free to focus their efforts on the isolation and characterization of the gene of interest. Second, the physical map provides a starting point for studying global genomic organization. As an increasing number of genes are cloned and molecular biological information is accumulated, one can begin to investigate the physical linkage of cloned genes, study the organization and distribution of repetitive elements, and address questions such as how physical

distance and genetic distance are correlated. In this context, the map provides the framework for cataloging and integrating molecular biological information. Ultimately, genome organization will be investigated at the nucleotide level. Clearly, physical maps are the logical substrates for genome-sequencing projects.

## Laborious Process

Physical mapping of complex genomes, however, is both laborious and computationally intensive. To illustrate the physical mapping problem, briefly described below are the researchers' efforts to assemble a complete physical map of the *Arabidopsis thaliana* genome, which will ultimately consist of a fully overlapping collection of cloned DNA fragments encompassing the five chromosomes.

The first stage of the mapping project involved the characterization of random cosmid clones by fingerprint analysis (Coulson et al, 1986; Hauge and Goodman, 1991). For the *Arabidopsis* project, approximately 20,000 random cosmid clones (~10 fold sampling redundancy) from primary libraries were fingerprinted. Using computer-matching programs, the clones have been aligned into some 750 overlapping groups or contigs. The contigs encompass approximately 90-95 percent of the *Arabidopsis* genome (Hauge et al, 1991).

In general, some 8-10 genomic equivalents must be fingerprinted to achieve 70-95 percent coverage of the respective genome. The task of ordering the clones and aligning them

with respect to the genetic map is formidable to say the least. To illustrate the magnitude of this problem, consider the maize genome, which is estimated to be 3,900,000 kb (kilobase). Using random cosmid clones containing an average insert size of 40 kb, approximately a million clones would be needed for 10 genomic equivalents.

## YAC Cloning Vectors

The mapping problem has been greatly simplified by the development of yeast artificial chromosome (YAC) cloning vectors (Burke et al, 1987). The YAC vectors allow for the routine cloning of 0.5 megabase-sized DNA fragments, representing an improvement of at least an order of magnitude over the previously existing techniques. The construction of YAC libraries involves the ligation of large DNA fragments (100-1000 kb) into a vector containing selectable markers and the functional components of a eucaryotic chromosome, autonomous replicating sequence elements for autonomous replication, the centromere for proper disjunction during meiosis and mitosis, and telomeres required for the replication of linear molecules (Murray and Szostak, 1983). The clones are transferred into baker's yeast (*S. cerevisiae*) where they are replicated along with the endogenous host chromosomes.

There are two clear advantages of the yeast-cloning system: The large size of the inserts means that fewer clones need be examined. Equally important is that YACs offer the potential to give a more random representation of clones than are obtained using conventional cloning systems.

## Utility of YAC Clones

The following examples of how YAC clones are being employed for *Arabidopsis* genome mapping illustrate the utility of YAC clones for genome research. Two general approaches are being used to assemble an overlapping YAC library covering the *Arabidopsis* genome. The first approach is to simply identify YAC clones corresponding to genetically mapped DNA probes (RFLPs and cloned genes). Presently some 380 *Arabidopsis* RFLP probes (Chang et al, 1988; Nam et al, 1989; S. Hanley and H.M. Goodman, unpublished; E. Meyerowitz, unpublished) are available for correlation of the physical map with the classical genetic linkage map. Using the available YAC libraries (Ward and Jen, 1990; Grill and Somerville, 1991), YAC clones corresponding to 125 RFLP markers have been identified (Hwang et al, 1991). Based on a mean YAC insert size of 160 kb and an average YAC contig size of 220-240, YACs of known genetic map location encompass approximately 30,000 kb or about 30 percent of the *Arabidopsis* genome (Hwang et al, 1991). Extension of this analysis to the remaining 160-some RFLP probes should result in a collection of YAC clones encompassing some 70 percent of the genome. Closure of the gaps can then be achieved by either chromosome walking or by utilizing the cosmid contig map as described below.

As an alternative strategy, the overlapping cosmid map (Hauge et al, 1991) is a powerful tool for assembling an overlapping YAC map. The linking strategy is to use the YAC clones to probe ordered arrays of cosmid clones that are

representative of the contigs (Coulson et al, 1988). Cosmids within a contig are chosen so that there is minimal overlap between flanking clones, yet the clones are representative of the contigs. The cosmids are plated onto nylon membranes as ordered arrays and subsequently probed with labeled YAC clones. Using this strategy, the gaps in the contig map are closed and the YACs are aligned within the framework of the cosmid map, thereby generating an overlapping YAC map.

An advantage of this approach over traditional strategies that use end-probes to select overlapping clones is that the hybridization patterns are easily tested for a logical fit to the structure of the contig map. Linkage can, therefore, be rapidly established based largely on the results of colony hybridization. In contrast, techniques such as genome walking require laborious confirmation of each join and subsequent restriction mapping of the linking clones to determine both the direction and the extent of the walk.

Using a combination of the techniques described above, it is probable that an overlapping YAC library of the *Arabidopsis* genome will be completed in the near future. The overlapping YAC library will serve to facilitate the cloning of genes and will provide a minimal set of clones covering the *Arabidopsis* genome. Given the small genome of *Arabidopsis*, a representative collection of clones can be gridded at high density onto a single filter the size of a microtiter dish. These "polytene" blots (J. Sulston and A. Coulson, personal communication) can then be

used to rapidly determine the chromosomal location of any new clone by simple blot hybridization.

YAC clones are likely to play an increasingly important role in future physical mapping projects. The strategies for physical mapping with YACs are essentially the same as those used for other genomic libraries (bacteriophage and cosmids). Using the existing technology, YAC clones may be fingerprinted directly and ordered into contigs (Kuspa et al, 1989). Moreover, the ability to easily generate endprobes from YACs using techniques such as inverse PCR (Ochman et al, 1988) allows for the construction of physical maps based on simple hybridization strategies.

The application of mapping strategies that use YACs should make it possible to undertake projects orders of magnitude larger than those currently underway. It remains to be determined, however, whether YACs will entirely supersede cosmid and  $\lambda$  clone maps since the smaller clones are generally required for routine procedures such as gene isolation and DNA sequencing.

## References

- Berg, D.E. and Howe, M.M. (1989). Mobile DNA. (American Society for Microbiology, Washington, D.C.).
- Burke, D.T., Carle, G.F., and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*, **236**, 806-812.
- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S. and Meyerowitz, E.M. (1988). Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, **85**, 6856-6860.
- Coulson, A., Sulston, J., Brenner, S., and Karn, J. (1986). Toward a physical map of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA*, **83**, 7821-7825.
- Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. *Nature*, **335**, 184-186.
- Feldman, K.A., Marks, M.D., Christianson, M.L., and Quatrano, R.S. (1989). A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science*, **243**, 1351-1354.
- Grill, E. and Somerville, C. (1991). Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Molec. Gen. Genet.*, in press.
- Hauge, B.M. and Goodman, H.M. (1991). Physical Mapping by Random Clone Fingerprint Analysis. In *Plant Genomes: Methods for Genetic and Physical Mapping*, T. Osborn and J.S. Beckmann eds. (Kluwer). In press.
- Hauge, B.M., Hanley, S., Giraudat, J., and Goodman, H.M. (1991). Mapping the *Arabidopsis* Genome. In *Molecular Biology of Plant Development*, G. Jenkins and W. Schurch eds. In press.
- Hwang, I., Kohchi, T., Hauge, B.M., Goodman, H.M., Schmidt, R., Cnops, G., Dean, C., Gibson, S., Iba, K., Lemieux, B.L., Danhoff, L., and Somerville, C. (1991). Identification and map position of YAC Clones comprising one third of the *Arabidopsis* genome. (submitted).
- Kohara, Y., Akiyama, K., and Isono, K. (1987). The physical map of the whole *E. coli* chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, **50**, 495-508.
- Kuspa, A., Vollrath, D., Cheng, Y., and Kaiser, D. (1989). Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA*, **86**, 8917-8921.
- Murray, A.W., and Szostak, J.W. (1983). Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.
- Nam, H.G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M., and Goodman, H.M. (1989). Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell*, **1**, 699-705.
- Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic Applications of an inverse polymerase chain reaction. *Genetics*, **120**, 621-623.
- Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., and Frank, T. (1986). Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA*, **83**, 7826-7830.
- Pruitt, R.E. and Meyerowitz, E.M. (1986). Characterization of the genome of *Arabidopsis thaliana*. *J. Mol. Biol.*, **187**, 169-183.
- Smith, C.L., Econome, J.G., Schutt, A., Klco, S., and Cantor, C.R. (1987). A physical map of the *Escherichia coli* K12 genome. *Science*, **236**, 1448-1453.
- Straus, D. and Ausubel, F.M. (1990). Genomic subtraction for cloning of DNA corresponding to deletion mutations. *Proc. Natn. Acad. Sci. USA*, **87**, 1889-1893.
- Van Sluys, M.A., Tempe, J., and Fedoroff, N. (1987). Studies on the introduction and mobility of the maize Activator element in *Arabidopsis* and *Daucus carota*. *EMBO J.*, **6**, 3881-3889.
- Ward, E.R. and Jen, G.C. (1990). Isolation of single-copy sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.*, **14**, 561-568. ♦

## Off the Wire



## Parser Available for GenBank® Flat File

Robert Read and Matthew Witten  
GenTools™ Project  
University of Texas Center for High Performance Computing

A software system available from the GenTools™ project at the University of Texas Center for High Performance Computing may be of interest to those who need to extract information from the GenBank® flat-file format.

The GenTools™ gbParse, the program parses GenBank® flat-file entries and translates them into a Prolog-like language. The software is expected to be useful to persons who cannot gain access to the (undoubtedly superior) relational format of GenBank® implemented in the RDBMS Sybase, or to those who wish to write special programs to extract information from the feature tables.

The parser software has been written using the UNIX and Free Software Foundation tools Flex and Yacc (or Bison). A C programmer can easily adapt the source code to produce output in any other required format.

The software is now in β-release. It has been tested on a SPARC Station and on a VAX/VMS system. Programmers might like to see the code (grammar) that has been written even if they do not intend to use it, as it represents the most concrete description of the GenBank® format, including the feature table.

Although the program translates 99% of the GenBank® entries, the code is not trouble free, in part because it must deal with actual syntax errors in the distributed flat files. GenTools™ gbParse has already been used to find numerous syntax errors in the distributed flat files. The program is robust in reporting entry errors.

To obtain the GenTools™ gbParse software and documentation, send an E-mail request to "gentools@chpc.utexas.edu" (Internet) or contact Robert Read, GenTools™ Project, UT-Center for High Performance Computing, Balcones Research Center, CMS 1.154, 10100 Burnet Road, Austin, Texas 78712. Further information about the GenTools™ project may be obtained from Dr. Sarah Barron at the same address. ♦

## Genome Sequencing Conference III Set for September

The annual Genome Sequencing Conference is an international conference devoted to discussion of the most current analyses and approaches to understanding the human genome. This year's conference will be held September 22-25 in Hilton Head, South Carolina, at the Hyatt Regency Hotel.

Innovative research by U.S., European, and Japanese groups will be presented and discussed. Presentations will detail discoveries from the human genome and model organisms, including mouse, *Drosophila*, *C. elegans*, yeast,

plants, *E. coli*, *M. capricolum*, and a number of viruses. State-of-the-art technologies and new computational approaches will also be covered. Poster sessions, workshops, and discussion groups will provide a forum for researchers to present their latest data. The conference co-chairs are J. Craig Venter (NIH) and Leroy Hood (CalTech).

Registration, which includes all meals and materials, is \$310 per person until August 2. After August 2, the cost is \$450. Students can register

for \$200 (no deadline), but a letter from their thesis advisor is required. The registration fee does not cover hotel expenses, but a room-sharing program is available to reduce the cost of the rooms. Early registration is encouraged. The deadline for receipt of abstracts is August 2.

For more information, contact Susan Wallace, P.O. Box 541, Rockville, MD 20848, Phone (301) 480-0634, FAX (301) 480-8588, E-mail: swallace@loglady.ninds.nih.gov. ♦

## Introducing Dr. Stephen Heller



Dr. Stephen Heller is the Informatics Project Leader for USDA's Plant Genome Research Program. Reporting directly to the

Program Director, Dr. Jerry Miksche, Dr. Heller manages the informatics portion of the program, primarily activities of cooperators and NAL staff. A major activity now underway is the development of a plant genome database system, which will provide users with genome data on four plant species--wheat, pine, corn, and soybean.

Dr. Heller has been a research scientist in USDA's Agricultural Research Service (ARS) since 1985. Before assuming his current position, Dr. Heller was a member of the Systems Research Laboratory. Responsibilities included developing and coordinating agency wide scientific database, modeling, and expert system programs; and developing a pesticide properties database and expert systems for evaluating analytical chemistry data.

From 1989 to 1990, on a leave of absence from ARS, Dr. Heller served as Director of Quality Control for Scitechinform, a UK-USSR joint scientific database venture of Maxwell Corporation and VINITI.

Before coming to ARS, from 1973 to 1983, Dr. Heller worked for the Environmental Protection Agency (EPA) as a project manager for a multi-agency, multi-organization,

NIH/EPA Chemical Information System (CIS), which served over 2,200 users in 22 countries. In 1976, in recognition of his contributions to the agency, he received EPA's Gold Medal.

Dr. Heller previously served as a senior staff fellow at the National Institutes of Health, from 1970 to 1973, and as a chemist for the U.S. Army from 1967 to 1969. During his career, he has served with the U.S. House Subcommittee on Health and the Environment (1979-80) and held the position of Lady Davis Visiting Professor of Chemistry at the Hebrew University in Jerusalem (1981).

He is an internationally well-known authority in scientific numeric and factual databases, and in chemical information. Dr. Heller is presently chairman of the IUPAC Committee on Chemical Databases and an American Chemical Society (ACS) COMP Division Councilor. He serves on the editorial boards of various journals, including the ACS Editorial Advisory Board for Computer Software. During the past 20 years, Dr. Heller has published over 130 papers and books, and has been the recipient of numerous fellowships and scholarships. He is a member of various professional organizations, including the ACS, IEEE, ASMS, CSR, AAAS, and Sigma Xi.

Dr. Heller received his B.S. degree in chemistry from the State University of New York, Stony Brook, and his Ph.D. in organic chemistry from Georgetown University in Washington, D.C. ♦

## Probe

ISSN: 1057-2600

The official quarterly publication of the USDA Plant Genome Research Program. This newsletter is aimed at facilitating interaction throughout the plant genome mapping community and beyond.

**Probe** is a publication of the Plant Genome Data and Information Center, National Agricultural Library.

Managing Editor  
**Susan McCarthy, Ph.D.**

Editor  
**Carolyn Bigwood**

Assistant Editor  
**Deborah Richardson, M.S.**

Production Manager  
**Terrance Henrichs**

Layout and Design  
**Barbara Binder**

Special Thanks to:  
**Dr. Michele Durand**  
**Dr. Ray Dolbert**  
**Dave W. Johnson**  
**Kathleen Hayes**  
**Dr. Rose Broome**

Articles, announcements, and suggestions are welcome.

**Correspondence Address**  
Susan McCarthy  
NAL, Room 1402  
10301 Baltimore Blvd  
Beltsville, MD 20705  
Phone:(301)344-3875  
FAX: (301)344-6098

**USDA Program Office**  
Dr. Jerome Miksche  
USDA/ARS/NPS/PNRS  
Room 331C, Bldg 005  
BARC-WEST  
Beltsville, MD 20705  
Phone:(301)344-2029  
FAX: (301)344-5467



## Beltsville Symposium XVI

Photomorphogenesis in Plants: Emerging Strategies for Crop Improvement.

September 22 - 26, 1991

University of Maryland University College Conference Center  
College Park, Maryland

Held jointly with the 1991 European Photomorphogenesis Symposium

### PROGRAM

Sunday	9/22/91	Registration and Evening Reception
Monday	a.m.	Session 1. Reflections on the Early Years of Phytochrome Research
		Session 2. Phytochrome Function in the Natural Environment
	p.m.	Session 3. Photoperiodism
	Evening	Posters
Tuesday	a.m.	Session 4. Blue Light Photoreception, Phototropism, and UV Photoreception
	p.m.	Session 5. Phytochrome Mutants and Transgenic Plants
	Evening	Posters
Wednesday	a.m.	Session 6. Phytochrome Regulated Gene Expression
	p.m.	Excursion
Thursday	a.m.	Session 7. Light Mediated Signal Transduction
	p.m.	Session 8. Phytochrome Structure and Properties
		Session 9. Opportunities and Future Research Directions
	Evening	Banquet

### Registration Fees

Regular (includes proceedings, reception, banquet)	\$175.00
Late Registration (postmarked after August 16, 1991)	\$225.00
Student (graduate & undergraduate; includes proceedings, reception, banquet)	\$75.00
(Other meals to be purchased on a cash basis)	

### Symposium Co-chairmen

William J. VanDerWoude and Steven J. Britz  
Climate Stress Laboratory, USDA, ARS  
Bldg. 046A, Beltsville Agricultural Research Center  
10300 Baltimore Blvd.  
Beltsville, Maryland 20705-2350

Telephone: (301) 344-3607  
Fax: (301) 344-4626  
Telex: 7402411 CLIM  
Bitnet: wvanderwoude@umdars

Sponsored by  
Beltsville Area, Agricultural Research Service  
in cooperation with  
Friends of Agricultural Research-Beltsville, Inc.

# Probe

Spring/Summer 1991

2 3

*fold*

Place  
Stamp  
Here

Plant Genome Data and Information Center  
USDA - National Agricultural Library  
10301 Baltimore Blvd., Room 1402  
Beltsville, Maryland 20705-2351

*fold*

Staple

*If you would like to continue receiving this publication, please return this page. If we do not receive a response within three issues your name will be removed from the mailing list.*

- ☐ Keep my name on the mailing list.
- ☐ Correct name on the label.
- ☐ Correct address on the label.
- ☐ Replace name on the label with my name.
- ☐ Add my name to the mailing list, but do not delete the name on the label.
- ☐ Remove my name from the mailing list.

///

///



Zip:

/ /

Telephone: ( ) - FAX: ( ) -

Photocopies of this slip are acceptable. Please submit entire page.

**Plant Genome Data  
and Information Center  
USDA - NAL  
10301 Baltimore Blvd.  
Beltsville, Maryland 20705-2351**

**BULK RATE**  
**POSTAGE AND FEES PAID**  
**USDA**  
**PERMIT NO. G95**